

語表付クローズテストの可能性

藤 枝 宏 壽

英語教室

(昭和56年11月16日 受理)

1 はじめに

クローズテストが、外国語としての英語 (EFL) の熟達度を総合的に測定する有効な用具であることは、すでに数多くの研究の実証するところであるが、⁽¹⁾ その高い関心の最大誘因は、このテストが “deceptively simple” ⁽²⁾ でありながら、信頼性・妥当性が極めて高いことにある。しかし EFL 学習者に対しては、原文より抜かれた語一本稿では「原語」という一のみを採点する exact word scoring method (EX) よりも、文脈上容認可能な語をも採る acceptable word scoring method (AC) の方が、有意差をもって勝れていることが明らかにされている ⁽³⁾ 以上、我々 EFL 担当教師は一つのジレンマに陥る。すなわち AC 法の厳密な容認判定には、英語を母国語とする者の援助が必要であり、ここに容易性というクローズテストの利点が減殺されるからである。

Jon Jonz⁽⁴⁾, Frances B. Hinofotis & Becky G. Snow⁽⁵⁾ および James D. Brown⁽⁶⁾ 等は、従来の ‘open-ended cloze’ を事前に試行し、その結果に基いて ‘multiple-choice cloze’ (MC) を構成した。そしてこの新クローズ実施の結果、高い信頼性と妥当性を検出し、採点の容易性の故に MC 有望なりと主張している。⁽⁷⁾ しかし、この MC にも難点がある。テスト作成段階で事前テストの分析という煩雑な作業が入ることと、次に述べる 4 肢選択形式のもつ限界である。

Jonz は、“The M-C cloze adds the feature of allowing the student a limited range of choices with which to compare the responses he himself generates……” ⁽⁸⁾ と論じている。しかしその ‘range’ が 25% の偶然性を孕む 4 肢選択であり、⁽⁹⁾ しかも既に語形変化をした語が提示されているのであるから、⁽¹⁰⁾ 学生は、時制・数など、文脈と重要な係わりを持つ文法的 productivity さえも奪われ、ただ項目毎に狭い視野で「消去法」によって正答を割り出すという再認作業に終始するであろう。従って、まず解答を ‘generate’ してみ、次にそれを 4 箇の選択肢と比較照合するという面倒な手順は、Jonz の理想論であって、学生に望むことは無理である。

そこで、MC クローズほどにテスト構成が煩雑でなく、採点も EX と同じ程度に容易であって、

しかも MC 以上に productive skills 測定の効力があり、信頼性・妥当性等は EX に勝り AC にも匹敵するクローズテストの形式が望まれることになる。

加瀬正二郎は、1979年、中学3年生に対して行った実験で、「補充される語を問題文の下に alphabet順で揚げたテスト（変形クローズ）」（空所は20）が、正規のクローズ（EX）に近い満足すべき信頼度係数（.766～.940）と、相当の併存的妥当性（.582～.662）を得たと報じている。⁽¹²⁾

筆者もすでに、この種の変形クローズが先述の要望を満す形式の一つであると予想し、語表に提示する語の数とその内容、および提示法に視点を置いて、以下に述べる2種類の語表付クローズテストの可能性を探ってみた。

2 実験計画

2-1 対象とテスト構成

高専2年生A・B2クラス、79名（1978年）に対して、クローズ法を利用した授業を2ヶ月間実施した。その過程で、H.A.Hillの“Elementary Stories for Reproduction”⁽¹³⁾から5篇（E9, E10, E16, E17, E29）と、“Intermediate Stories for Reproduction”⁽¹⁴⁾から2篇（I3, I5）とを用いて、語表付クローズテストに関する一連の試行的小実験を行なった。

上記のテスト材料は、いずれも約150語で書かれた小話で、FleschのReadability Score⁽¹⁵⁾によると、‘very easy’から‘fairly easy’の間にランクされ、高専2年生にとっては比較的易しい教材である。これを次の3形式・4種のクローズテストに構成した。（付録参照）

- (1) 標準クローズ(EX, AC) —— 6語目毎の単語を抜いて25の空所を作ったもの。これを厳正法(EX)と容認法(AC)で採点する。
- (2) 原語表付クローズ(WL1) —— 空所作成は(1)と同一。これに、補充されるべき語の全てを、inflectionやcontractionがあった場合には原形に戻し、アルファベット順に配列した「原語表」を付ける。同一語が2回以上使用されることもあるので、実際に提示したのは20～21語である。
- (3) 増語表付クローズ(WL2) —— 空所作成は(1)と同一。ただし、(2)の原語表に、さらに錯乱肢(distractors)を加えた50語の「増語表」を付ける。錯乱肢については、その約2/3は同品詞から、残りの約1/3は異品詞から、文脈上正答に近似しないものを教師が選定した。語の配列、提示は(2)と同じである。

採点は、WL1もWL2もEX法で行なう。

2-2 実験の目的

語表付クローズテストの可能性

上記2種の語表を作成した意図は2つある。第1は、20～50語の語表付加によって、1空所に対する適語選択のrangeが4肢選択の場合に比べて遙かに大きくなり、学生の受動的消去作業の可能性が減じ、また偶然性介入の危険も少くなる一特に増語表においてその効果が期待されるということである。第2には、表中の語が全て原形で提示されるため、学生が語尾変化や縮約を行なう必要も生じ、それによって、文脈的・文法的productivityが増加することを目論むものである。

この意図の下に3形式・4種のクローズの比較実験を行なうに当り、次の仮説を立てた。

H1: WL1およびWL2は、EX以上に、そしてACと同程度に有効なEFL総合学力の測定用具である。

H2: WL2はWL1よりも有効なEFL総合学力の測定用具である。

H3: 語表の原形提示はテストの弁別力を増す。

2-3 実験のデザイン (表1 参照)

実験I: まず対象2クラスの等質検定をするため、共通の問題で4種のクローズテスト (T₂～T₅)を行なう。また、等質検定のためと、妥当性の外部規準とするため、読解・文法作文・LLを含む2年英語学年成績「英語II」を別途算出し、T₁とする。

実験II: 実験IのT₂～T₅のデータをクローズ4種の特性観察に使用する。

実験III: WL1とEX/ACとの比較、およびWL2とEX/ACとの比較を、それぞれ2回ずつのテスト (T₆, T₇およびT₈, T₉)により、毎回クラスを交替させて行なう。どのテストも制限時間は25分とする。

このように同趣の実験を繰返すのは、テスト材料、テスト構成、標本等の偏りがなるべく丸められて、目指す語表の効果測定が少しでも一般化されることを狙うためである。

(表1) 実験のデザイン

テ ス ト			クローズ4種の割当		実 験 目 的	実 験 番 号
No.	材 料	難 易 度	Aクラス(N=42)	Bクラス(N=37)		
T ₁	英語II		(2年英語総合成績)		等質検定	実験I
T ₂	E9	very easy	EX		等質検定	
T ₃	E9	very easy	AC		および	
T ₄	I3	fairly easy	WL1		EX, AC, WL1,	実験II
T ₅	E10	very easy	WL2		WL2の特性観察	
T ₆	E17	easy	WL1	EX, AC	WL1とEX/AC	実験III
T ₇	I5	easy	EX, AC	WL1	との比較	
T ₈	E29	easy	WL2	EX, AC	WL2とEX/AC	
T ₉	E16	easy	EX, AC	WL2	との比較	

3 実験の結果と検討

3-1 統計処理

クローズテストの採点は、項目分析を含めて全て2回ずつ行ない、それぞれ得点(25点満点)のクラス平均(M), 標準偏差(SD), キューダ・リチャードソン公式20による信頼係数(KR), 外部規準「英語II」との相関係数(r), 項目分析による項目通過率(IF)と有意弁別項目数(DI)を算出した。また必要に応じて、Mの差, r , および r の差についてt-検定を, SDの差についてはF-検定を行なった。以下, その結果を実験デザインの順に検定する。

3-2 実験I・等質検定(表2参照)

T₁ 英語II, およびT₂~T₅いずれの形式のクローズにおいても, A・B両クラス間に得点の有意差は認められなかった。標準偏差も, T₂~T₅においては両クラス間に有意差がないが, T₁の場合だけ5%レベルで有意差が生じ, Aクラスの散布の方が大きいことを示している。この広がり差は平生の授業における観察と一致しており, その事は,

(表2) 対象2クラスの等質検定

T₂~T₅の各テストとT₁との相関係数において, 有意差こそないものの, 総じてAクラスの方がやや低いことと関連があるようである。この点に注意さえすれば, A・B両クラスは以下のクローズ4種の比較実験に耐えうる標本であると判定した。

		Aクラス	Bクラス	A-B
T ₁ ・英語II	M	73.6	75.7	-2.1
	SD	10.8	7.6	3.2*
T ₂ ・EX	M	11.4	10.9	0.5
	SD	2.2	2.5	-0.3
	r	0.49**	0.46**	0.03
T ₃ ・AC	M	14.2	13.7	0.5
	SD	3.0	3.5	-0.5
	r	0.52**	0.54**	-0.02
T ₄ ・WL1	M	12.9	12.3	0.6
	SD	2.9	2.7	0.2
	r	0.58**	0.61**	-0.03
T ₅ ・WL2	M	11.6	11.0	0.6
	SD	3.3	3.0	0.3
	r	0.39*	0.63**	-0.27

(*印: $p < 0.05$ で有意 **印: $p < 0.01$ で有意)

3-3 実験II・クローズ4種の特性観察(表3参照)

クラス別に統計した表2のT₂~T₅を, 両クラス合わせて1標本と見なして集計したのが表3である。T₂とT₃は同一テストであるが, T₄とT₅はそれぞれ別のテスト材料を用いてあるので, T₂~T₅の厳密な比較はできない。しかし同じ著者による同趣のテスト文であるから, EX, AC, WL1, WL2のクローズ4種について一応の特性観察は可能であろう。

まずEXとACを比較すると, 平均点(M)から弁別項目数(DI)までの全ての測定においてACが勝れていることは, 従来の諸研究の結果と一致している。特にその信頼係数(KR) .79は抜群に高く, 弁別項目数も最多であって, ACをWL1, WL2の到達目標とすることは妥当と思われる。

る。EXは'very easy'級のテスト文でありながら、なお通過率が悪く、そのM10.9はクローズ4種中最低であり、散布度もSD 2.4と萎縮している。したがって信頼係数、相関係数、弁別力も最低となり、EFL学生にEXクローズを用いることにかなりの問題があることを暗示している。

WL1, WL2の得点12.3, 11.4はACの13.7には及ばないが、EXの10.9よりも高いことから、語表をcueとし

て与えた一応の効果は認められる。しかしWL1のテスト文が中で最も難しい'fairly easy'のランクにあることを考慮すると、他と同一レベルのテスト文であれば、その得点はさらに伸びるものと推測される。

標準偏差、信頼係数、相関係数、弁別項目数のいずれにおいても、WL1, WL2は共にACとEXの間に位置しているが、そのKRの値.51, .53およびSD2.7, 3.0がACのKR.79, SD3.5に比べて相当に低いのが問題である。それがテスト文の内容や、テスト構成の差異に起因するのか、それとも語表付加そのものの影響であるのか、その結論は後の実験結果を俟たねばならない。

相関係数(r)では4種間に有意差はないが、いずれも外部規準による妥当性係数の有意基準.45⁽⁶⁾は超えており、英語総合学力測定の妥当性は、十分に高くはないが、一応認められる。

以上、この段階での観察をまとめれば、WL1とWL2との間に明確な優劣はつけ難いが、テスト特性に重要なkeyとなるSD, KRの観点から、WL2がやや優位にある。したがって、総合的テスト特性は、 $AC > WL2 \geq WL1 > EX$ と推定できる。

3-4 実験Ⅲ・交替法によるEX/ACとWL1, EX/ACとWL2の比較(表4参照)

得点状況(M)から調べる。 $T_6 \sim T_9$ を通じ、 $AC > EX$ の型は1%レベルで有意であり、前節における観察と一致する。WL1, WL2は、EXに対してはいずれも1%レベルの有意差をもって高いが、ACに対しては T_6 のWL1の場合を除き有意差がない。したがって、増語表を付加すれば少くともAC程度の通過率を得、原語表を付加すればACと同程度か、それ以上の通過率を得ることができる。と云える。

標準偏差(SD)については、 T_9 のWL2 $>$ EXの場合を除いて、どのテストにおいても各クローズ3種間に有意差はない。しかし、名目数値を比較すると、 $WL1 > AC \geq EX$, $WL2 \div AC > EX$ の傾向が読み取れる。語表付加によって少くともAC程度に散布が開くということは、重要な特性として注目すべきである。

信頼係数(KR)においても、クローズ4種間の関係は上記SDの場合とはほぼ同様の傾向を示す。

(表3) クローズ4種の特性観察

	$T_2 \cdot EX$	$T_3 \cdot AC$	$T_4 \cdot WL1$	$T_5 \cdot WL2$
M	10.9	13.7	12.3	11.4
SD	2.4	3.5	2.7	3.0
KR	.47	.79	.51	.53
r	.49**	.54**	.54**	.53**
IF	.44	.56	.49	.46
DI	11	16	14	13

(**印: $p < 0.01$ で有意)

(表 4) 交替法による EX/AC と WL 1, EX/AC と WL 2 の比較

	T ₆			T ₇			T ₈			T ₉		
	EX	AC	WL1	EX	AC	WL1	EX	AC	WL2	EX	AC	WL2
M	11.6	14.5	18.6	10.0	13.3	12.9	10.1	11.9	12.6	7.8	10.7	10.6
SD	3.5	3.7	3.9	2.2	2.2	3.0	3.1	3.5	3.3	2.3	3.2	3.5
KR	.74	.74	.76	.41	.33	.51	.60	.66	.64	.43	.63	.68
r	.50**	.66**	.40**	.23	.43**	.46**	.59**	.62**	.37**	.48**	.52**	.51**
IF	.46	.58	.74	.40	.53	.52	.40	.48	.50	.31	.43	.42
DI	12	11	15	7	3	5	8	8	10	4	6	10

すなわち、T₇においては3種ともKR値の低いのが問題として残るが、T₆、T₇においてWL1がEX、ACよりも高く (WL1 > EX ≥ AC)、また、T₈、T₉においてWL2がEXを超え、ACにほぼ等しいか、さらにそれを凌ぐほどに信頼性を増している (WL2 ≥ AC > EX)。

上記SD、KRにおけるWL1、WL2の有望な結果から判断すれば、前節で提起した「語表付加によってSD、KRが低下するか」という疑念は捨てられてよからう。テスト構成に注意すれば、語表付加はかえってSD、KRを高めるものと云える。

相関係数 (r)、すなわち妥当性係数については、T₇・EXの .23を除き、全てのテストにおいて、 r そのものの有意性は1%レベルで検定されたが、各テストのクローズ3種間に r の有意差は検出されなかった。しかし、名目数値で比較すると、ここでもAC > EXの傾向は各テストに共通して認められるが、EX/ACとWL1またはWL2の対比においては、傾向が大きく二分されていることに気付く。すなわち、語表側が最低となるT₆(AC > EX > WL1)およびT₈(AC > EX > WL2)の場合と、逆に語表側が高いT₇(WL1 > AC > EX)およびT₉(AC ≅ WL2 > EX)の場合の対立である。この矛盾を解くのはAクラスである。T₆、T₈とともに語表側にまわっているのがAクラスであり、そのAクラスは、先述の等質検定で、Bクラスよりも r が低くなる傾向をもつと指摘された。その傾向が今、実験Ⅲにおいても現われたものと解釈できよう。しかも、妥当性係数の有意基準 .45を下廻っているのは、そのT₆・WL1の .40、T₈・WL2の .37 と、T₇・EXの .23および同ACの .43の4回だけであり、いずれもAクラスの英語Ⅱの分布の特異性に起因したものと推測される。ただし、T₇のテスト特性が全般的に劣っていることは否めない事実であり、これはテスト構成の研究課題である。

したがって、EX/AC との対比におけるWL1、WL2の外部規準妥当性については、実験Ⅲの結果だけでは即断できない。実験Ⅱの結果を援用して一応の結論を出せば、WL1およびWL2の外部的妥当性は、少なくともEX以上はあり、ACに近づく傾向にあると推定される。

3-5 項目分析 (表5, 表6 参照)

統計的妥当性を検討する今一つの方法は、内部的規準—問題項目の困難度、すなわち通過率 (IF) と、項目弁別力 (ID) によるものである。⁽¹⁷⁾ 実験Ⅲの各々のテストについては、表4のIFおよびDIの欄にその結果を示した。また、WL1に係わるT₆, T₇と、WL2に係わるT₈, T₉を、それぞれ1つのテスト ((T₆₇, T₈₉) とみなして集計し、機能語(F)と内容語(C)別に⁽¹⁸⁾ IF, IDを百分率でそれぞれ表5, 表6に示した。なお、比較を容易にするため、各EXの「全項目」の百分率値を100とした場合の指数を括弧内に示してある。

項目通過率から検討する。表5のT₆₇, T₈₉・「全項目」の指数によりクローズ4種を比較する

(表5) クローズ4種の機能語・内容語別項目通過率 (IF) %と(指数)

	T ₆₇			T ₈₉		
	EX	AC	WL1	EX	AC	WL2
機能語(F)	57.1(133)	63.3(148)	70.1(164)	46.9(132)	54.1(152)	53.1(149)
内容語(C)	35.2(82)	51.7(121)	60.9(142)	28.0(79)	39.6(111)	42.8(120)
全項目	42.8(100)	55.4(129)	63.4(148)	35.6(100)	45.4(128)	46.9(132)
F-C	21.9(51)	11.6(27)	9.2(18)	18.9(53)	14.5(41)	10.3(29)

と、およそWL1>WL2>AC>EXの関係が認められ、3-4での検討結果とほぼ同じである。しかし、通過率については.50を理想とする考えがある。⁽¹⁹⁾ その基準に近いものから配列すれば、AC(55.4), EX(42.8), WL1(63.4) および WL2(46.9), AC(45.4), EX(35.6) となる。したがってクローズ4種の通過率の評価は、WL2>AC>EX>WL1とまとめられよう。

次に機能語と内容語に分けて語表付加の意義を考える。両語類の差(F-C)の欄を見ると、クローズ4種いずれにおいても機能語の方が通過率が高いが、⁽²⁰⁾ その差がEXにおいて最も大きく、AC, WL2, WL1の順に減少していることに注意すべきである。すなわち、語表付加の効果は、単に全体の通過率を上げること以上に、機能語よりも内容語の正答を容易にし、その効果はWL2よりも、提示語数の少いWL1において強く現われるのである。換言すれば、語表付加による適語選定の範囲限定という作用が、“closed-class items”といわれる機能語よりも、“open-class items”といわれる⁽²¹⁾ 内容語に強く働くことを示している。

問題項目の弁別力について、まず表4・DIの欄を見ると、T₇を除きWL1, WL2の弁別力がEX/ACに勝っていることが分る。これを表6の指数を利用してさらに比較検討する。全項目においては、WL2(167)が抜群に高いが、WL1(105)は案外伸びておらず、EX(100)に近い。また、T₆₇のAC(74)が、これまでのデーター (AC>EX) ⁽²²⁾ に反して、大きく逆転している。これはT₇のテスト構成に問題があり、容認語の採点によって下位群の成績が向上し、EXの時の弁別項目数の6割近くが失われてしまった為である。(表4のDIの欄参照)

(表 6) クローズ 4 種の機能語・内容語別項目弁別力 (ID) %と(指数)

	T ₆₇			T ₈₉		
	EX	AC	WL 1	EX	AC	WL 2
機能語(F)	44.4(117)	33.3(88)	5.6(15)	20.0(83)	15.0(63)	35.0(146)
内容語(C)	34.4(91)	25.0(66)	59.4(156)	26.7(111)	36.7(153)	43.4(181)
全 項 目	38.0(100)	28.0(74)	40.0(105)	24.0(100)	28.0(117)	40.0(167)
F - C	10.0(26)	8.3(22)	-53.8(-141)	-6.7(-28)	-21.7(-90)	-8.4(-35)

機能語と内容語との対比における語表付加の弁別効果は一様ではない。T₆₇のEX, AC は共に F > C の型を示しているのに、WL1は逆に F < C の傾向を強く示している。これは、T₇の問題もさることながら、T₆においても、その弁別項目数対該当項目合計数の比が、内容語で15/18、機能語で0/7 という極端な偏りを生じていることに原因がある。すなわち、21語の原語表の中で、機能語は5語しか含まれていなかったために、その中から適語を選ぶことは極めて容易となり一現に表5のT₆₇・WL1の機能語の通過率は70.1(164)と表中最高であって一上位・下位両群間に得点の有意差が生じなかったためである。この点、原語表付クローズテスト構成の際、空所の数、機能語・内容語の比率等を慎重に考慮すべきことを示唆するものといえよう。

T₈₉のWL2においては、EX, ACと同様に C > F の型を示しており、F/C別の弁別効果に関する限り、WL2の特殊性は見られない。

以上の検討結果から判断すれば、実験Ⅲにおける語表付加の弁別効果は、WL1は僅かに、WL2においては顕著に、EX, ACを上廻るものといえる。

3-6 原形提示の効果 (表7 参照)

WL1, WL2の語表には、全ての語を語尾変化を伴っていない原形で提示したが、その効果を表7によって検討する。表7は、T₆~T₉各々の問題項目全体を、語尾変化・縮約を必要とする項目(I)と、それを必要としない項目(R)とに二分し、それぞれについて通過率と項目弁別率(弁別項目数/該当項目合計数)とを算出し、前節と同じくT₆₇, T₈₉の2群にまとめて表示したものである。

まずクローズ4種の通過率をI/R別に比較すると、次のようになる。

I : AC > WL1 > EX, AC > WL2 > EX

R : WL1 > AC > EX, WL2 > AC > EX

R項目においては、WL1もWL2もEX, ACを凌ぐほどに語表付加のcue効果が端的に現われている。他方、I項目においては、WL1,特にWL2において困難度が増加し、ACを下廻る²³⁾通過率となっているが、EXほどには悪くない。因みに、IのRに対する比率をWL1, WL2について試算すると、それぞれ.95, .49となり、原形提示がWL2においては相当に厳しい負荷となっ

ていることが分る。

その負荷の質を項目弁別力について調べると、次の関係が成り立つ。

$$I : WL1 > EX > AC$$

$$WL2 > EX = AC$$

$$R : EX > AC > WL1$$

$$WL2 > AC > EX$$

I 項目において原形提示の項目弁別効果が明らかに見られ、特に W

L2は大差をもってEX, ACに勝っている。一方、R 項目においてWL1とWL2の位置の逆転が目立つが、R 項目各3 種間の弁別率の最大差(6.1, 7.9)は、I 項目のそれ(41.2, 41.6)と比べれば僅少である。I とR の該当項目合計数はそれぞれ、T₆₇において17と33、T₈₉において12と38であるから、弁別力の総量においては依然としてR 項目の方が多いが、弁別の効率からいえばI 項目が多分に優勢である。

かくして、語表に原形提示という手法を用いたことは、安易なcue 効果を抑えて文脈的・文法的productivityを要求し、それによって内的弁別力を高めたという点で、有効であったといえる。

4 結論

(表8) 仮説および実験結果の図式

仮説H1	$AC \cong (WL1, WL2) > EX$
H2	$WL2 > WL1$
H3	$I > R^*$
実験II	$AC > WL2 \geq WL1 > EX$
実験III S D	$WL1 > AC \geq EX, WL2 \cong AC > EX$
K R	$WL1 > EX \geq AC, WL2 \geq AC > EX$
r	$AC > EX > WL1, AC > EX > WL2$
	$WL1 > AC > EX, AC \cong WL2 > EX$
	$\rightarrow \{AC > (WL1, WL2) \geq EX\}$
I F	$AC > EX > WL1, WL2 > AC > EX$
	$\rightarrow \{WL2 > AC > EX > WL1\}$
I D	$WL1 > EX > AC, WL2 > AC > EX$
	$\rightarrow \{WL2 > WL1 > (AC, EX)\}$
I/R	$I > R^*$

(* I = 原形の語尾変化等を要する項目の弁別力
R = 原形のまゝ補充できる項目の弁別力)

(表7) 原形提示の効果：語尾変化必要項目(I)・不必要項目(R)別通過率と弁別率 %

		T ₆₇			T ₈₉		
		EX	AC	WL1	EX	AC	WL2
通過率	I	46.7	68.5	60.8	18.9	39.7	25.7
	R	42.6	51.1	64.1	41.2	46.9	52.8
弁別率	I	41.2	17.6	58.8	8.4	8.4	50.0
	R	36.4	33.3	30.3	28.9	34.2	36.8

前章では、実験結果について測定項目毎にクローズ4 種を比較検討したが、最後にこれを総括して、仮説に対する結論を下したい。全体の把握を容易にするため、仮説および実験結果を図式⁽²⁴⁾によって整理し、表8に示す。

実験IIの結果はH1承認の方向にあり、また実験IIIの諸データもH1を支持し、さらにH1を超えてWL2, WL1を評価するものが多い。H2についても、実験IIおよび実験IIIのIF, IDは、その成立を認める資料である。H3は前節3-6ですでに証明されている。

したがって、今回の実験に限っていえば、原形提示法を用いた語表付クローズテスト、特に増語表付クローズテストは、標準クローズの厳正法(EX)以上に有効であり、容認法(AC)に勝るとも劣らぬ有効

なEFL学力測定用具として、多分にその可能性を有するものであると結論できる。

しかし、未解決の問題がいくつか残っている。第一には、今回の小実験において外部的妥当性の測定が不十分であったことである。それは、標本の数に限られていたこと、proficiencyというよりもむしろachievementの評価である「英語II」を外部規準としたこと、それに、英語IIの成績散布がクラスによって偏りが生じていたこと等が原因であったと思われるが、この点は追検を要する課題である。次に、WL1よりもWL2の弁別力が勝っていたのは、20～21対50という提示語数の差異によるものか、それとも後者に混入した錯乱肢の効果であるのかという問題もある。語表付クローズの空所の適数、錯乱肢の選定法とともに研究されるべきである。今一つ、被験者のproficiencyのレベルによるWL、WL2の適合度の問題も検討されていない。本稿では、高専2年生（高校2年レベル）を対象とした実験だけを扱っている。初級や上級の場合に、二種の語表付クローズがどのような効力を示すかということは、興味ある問題である。

クローズテストが常に万能であるとは勿論いえないが、簡便に、しかも正確にEFLのproficiencyを測定しなければならない場面においては、大いに利用価値がある。その簡便さと正確さを兼備するものとして、語表付クローズの可能性が今後さらに追究されることを期待するものである。

注

- 1) Frances Butler Hinofotis, "Cloze as an Alternative Method of ESL Placement and Proficiency Testing," (ed.) John W. Oller, Jr. & Kyle Perkins, *Research in Language Testing*, Newbury House, 1980, p. 122 参照。
- 2) John W. Oller, Jr., "Cloze Tests of Second Language Proficiency and What They Measure," *Language Learning*, 23(1973), p. 105 参照。
- 3) John W. Oller, Jr., "Scoring Methods and Difficulty Levels for Cloze Tests of Proficiency in English as a Second Language," *Modern Language Journal*, 56(1972), p. 157 参照。
- 4) Jon Jonz, "Improving on the Basic Egg: The M-C Cloze," *Language Learning*, 26(1976), pp. 255-265 参照。
- 5) Frances Butler Hinofotis & Becky Gerlach Snow, "An Alternative Cloze Testing Procedure: Multiple-Choice Format," (ed.) John W. Oller, Jr. & Kyle Perkins, *Research in Language Testing*, Newbury House, 1980, pp. 129-133 参照。
- 6) James Dean Brown, "Relative Merits of Four Methods for Scoring Cloze Tests," *Modern Language Journal*, 64(1980), pp. 311-317 参照。
- 7) 特に F.B. Hinofotis & B.G. Snow (前掲書, p. 129) と J. Jonz (前掲書 p. 263) の主張。

語表付クローズテストの可能性

- 8) Jon Jonz, 前掲書p. 256参照。
- 9) このことについてはF.B.Hinofotis & B.G.Snow (前掲書p. 133) も認めている。
- 10) J.Jonz (前掲書p. 259) の用いたM-Cクローズの例を参照。
- 一例: ……Even if $\frac{14}{14}$ only nine o'clock by the cathedral clock,
I'm going right to bed." 14-is it's the you
- 11) いわゆる「消去法」は、幼少から多肢選択テストに慣らされている学生の常識である。
- 12) 加瀬正二郎「クローズテストと変形クローズテストとの比較について」『中部地区英語教育学会紀要』9 (1979), 107-116頁参照。
- 13) L.A.Hill, *Elementary Stories for Reproduction*, Oxford University Press, 1971.
- 14) L.A.Hill, *Intermediate Stories for Reproduction*, Oxford University Press, 1971.
- 15) Rudolf Flesch, "A New Readability Yardstick," *Journal of Applied Psychology*, 32 (1948), pp.221-233参照。
- 16) 橋本重治『教育評価法総説』金子書房 (昭和38年) 105頁参照。
- 17) 上掲書104-106頁参照。
- 18) Randolph Quirk & Sidney Greenbaum, *A University Grammar of English*, Longman, 1973, p.18参照。
- 19) 橋本重治 前掲書106頁, J.D.Brown 前掲書p. 315参照。
- 20) John W.Oller, Jr. & Nevin Inal, "A Cloze Test of English Prepositions," *TESOL Quarterly*, 5 (1971), p. 315参照。W.L. Taylor (1956) が同じような結果を報告しているとの言及がある。
- また、加瀬正二郎 (前掲書 111頁) も同様の結果を報じている。
- 21) R.Quirk & S. Greenbaum前掲書p. 316参照。
- 22) J.D.Brown 前掲書p. 315; 藤枝宏壽「Cloze Testについての一考察」『福井工業高等専門学校研究紀要 人文・社会科学』10号 (1977) 42頁参照。
- 23) 語尾変化等を要する I 項目で、通過率がAC>WL2となったのは、語尾変化の点を甘く容認(AC)したからというのではなく、偶々文脈上容認できる語がいくつかI 項目の中にあっただからである。その場合も勿論、語尾変化等、文法的に正しいものを容認した。
- 24) この図式は名目数値によって比較した結果を記したものである。及ずしも有意差ではない。

付録

1) WL1の例：T₆ (E17)

☆次の笑い話は、6語目ごとに単語が脱落しています。文脈や文法をよく考えて各空所に適当な1語を補いなさい。(制限時間25分)

It was Jimmy's birthday, and he was five years old. He ¹(got) quite a lot of nice ²(birthbay) presents from his family, and ³(one) of them was a beautiful ⁴(big) drum

"Who gave him that ⁵(thing)?" Jimmy's father said when he ⁶(saw) it.

"His grandfather did," answered ⁷(Jimmy's) mother.

"Oh," said his father.

⁸(Of) course, Jimmy liked his drum ⁹(very) much. He made a terrible ¹⁰(noise) with it, but his mother ¹¹(did) not mind. His father was ¹²(working) during the day, and Jimmy ¹³(was) in bed when he got ¹⁴(home) in the evening, so he ¹⁵(did) not hear the noise.

But ¹⁶(one) of the neighbours did not ¹⁷(like) the noise at all, so ¹⁸(one) morning a few days later, ¹⁹(she) took a sharp knife and ²⁰(went) to Jimmy's house while he ²¹(was) hitting his drum. She said ²²(to) him, "Hullo, Jimmy. Do you ²³(know), there's something very nice inside ²⁴(your) drum? Here's a knife. Open ²⁵(the) drum and let's find it."

Word List (1)

be	big	birthbay	do	get	go
home	Jimmy	know	like	noise	of
one	see	she	the	thing	to
very	work	you			

備考：左の表の単語は残らず
使用されます。

注意：上の単語表から適当な1語を選んで、問題文の空所を補いなさい。ただし次のような場合がありうるから注意しなさい。

- (1) 同じ語が2回以上使われる場合。
- (2) 適当に語形を変える必要のある場合。例：boy→boys, boy's, boys' ; I→my, me (格や数の変化) / tall→taller, tallest (比較変化) / be→am, are, is, was, were, been, being (動詞の活用)
- (3) 2語が縮約されて1語となる場合。例：do+not→don't ; I+am→I'm.

語表付クローズテストの可能性

2) WL2の例：T₀ (E16) (冒頭の問題指示文はE17と同一)

John liked chocolates very much, but his mother never ¹(gave) him any, because they were ²(bad) for his teeth, she thought. ³(But) John had a very nice ⁴(grandfather).

The old man loved his ⁵(grandson) very much, and sometimes he ⁶(brought) John chocolates when he came ⁷(to) visit him.

Then his mother ⁸(let) him eat them, because she ⁹(wanted) to make the old man ¹⁰(happy).

One evening, a few days ¹¹(before) John's seventh birthday, he was ¹²(saying) his prayers in his bedroom ¹³(before) he went to bed. "Please, ¹⁴(God)," he shouted, "make them give ¹⁵(me) a big box of chocolates ¹⁶(for) my birthday on Saturday."

His ¹⁷(mother) was in the kitchen, but ¹⁸(she) heard the small boy shouting ¹⁹(and) went into his bedroom quickly.

"²⁰(Why) are you shouting, John?" she ²¹(asked) her son. "God can hear ²²(you) when you talk quietly."

"I ²³(know)," answered the clever boy with ²⁴(a) smile, "but Grandfather's in the ²⁵(next) room, and he can't."

Word List (2)

a	after	although	and	angry	ask
bad	be	before	big	boy	bring
but	can	come	do	father	for
from	give	God	grandfather	grandson	happy
have	he	help	I	just	know
let	mother	next	on	or	say
she	sir	strong	take	tell	the
to	uncle	very	want	who	why
write	you				

(注意書はE17と同一)

Potentialities of Word-List Cloze Tests

Koju FUJIEDA

Of the various scoring methods for the cloze test, the acceptable word scoring method (AC) has been advocated as the most effective in EFL situations. However, contrary to the reputed simplicity of a cloze procedure, the AC method has a drawback to EFL teachers in that it requires a glossary of acceptable answers to be established from the native-speaker sample.

The multiple-choice cloze test may be a feasible solution to this problem, but it also needs laborious pretest work to construct powerful test items. Moreover, this modified cloze test seems to be too much involved in measuring receptive skills rather than productive skills.

The author conducted on 79 technical college students a series of experiments with two types of 'word-list cloze test,' the one with a list of the deleted words (WL1) and the other with a list of the deleted words plus distractors (WL2). All the words listed were arranged in the alphabetical order and presented in their root forms without inflection and/or contraction.

The hypotheses for these experiments were:

- 1 As a test of measuring overall EFL proficiency, WL1 and WL2 are more effective than the exact word scoring method and as effective as the AC.
- 2 As an overall EFL proficiency test, WL2 is more effective than WL1.
- 3 The root-form presentation of alternatives is more powerful in discrimination than otherwise.

The statistics, computed and examined on the basis of the mean score, standard deviation, reliability coefficient (K-R20), correlation coefficient with English II, item facility and item discrimination, showed that all the hypotheses could be adopted on condition that the matter of criterion-related validity be explored further. The word lists were found to facilitate the filling-in of content words more than function words.

Potentialities of word-list cloze tests are now open to be tapped for practical use.